ISDN Personal Video

Paul E. Crouch
J. Al Hicks
John J. Jetzt

Desktop personal video telephony has become an area of increased interest among telecommunications professionals. This interest has been spurred by the combination of increased availability and lower cost for the bandwidth required for video transmission, and significant advances in video compression technology. Although analog technology current supports video telephony, the Integrated Services Digital Network (ISDN) will increase the growth and development of video telecommunications.

Introduction

ISDN's performance qualities make it an excellent and effectively noise-free transmission medium for voice, data, and video. Its error rate is on the order of 10^{-7} errors per transmitted bit. In addition, many ISDN interfacing arrangements have been designed to allow clear-channel communications without any processing in the network such as level changes and echo control.

Thus, ISDN provides a major advantage in terms of quality and bandwidth when compared to today's analog interface arrangements. Although it has been initially expensive, as ISDN is increasingly adopted, it will become more and more economical to operate through economies of scale. Conformity with international standards makes ISDN an even more logical choice for high quality telephone transmission.

Today the telecommunications community is thinking positively about desktop personal video telephony: real-time, two-way video and audio communication involving two or more people. Although video telephony has long been technically possible and even seriously proposed, until lately several major problems prevented its introduction, the primary one being that the bandwidth (i.e., data transmission rate) required for video transmission was prohibitively expensive and was not generally available.

Today, technology has met this challenge. Wideband transmission has become less expensive at the same time that video compression techniques have reduced the bandwidth required for video. As a result, the formerly "futuristic" video telephone is becoming a reality. As the technical and economic feasibility of videotelephony is proven, it will stop being "the wave of the future" and will move instead into the present. All signs suggest that the remainder of the Nineties will be a "video decade" where personal video will become commonplace.

We have already seen the introduction of a practical analog video telephone, the AT&T Videophone 2500, which uses the existing switched analog telephone network for its connectivity. However, significant advantages for personal video telephony can be gained by using ISDN.

Video Compression Technology and ISDN

Video compression technology has significantly reduced the bandwidth needed for video transmission. Moreover, as exemplified by the Videophone 2500, video compression technology is ready for practical, printed circuit board-sized applications that can be a part of a practical video telephone. Also, because of evolving compression standards, video compression technology also is becoming compatible with different vendors' video telephones.

High quality, economically priced cameras and displays for personal-sized video telephones are becoming available. Hence, video telephony is benefiting from the growth of demand in the consumer electronics market, i.e., from the cameras used in video camcorders, and the displays used in electronic games, and miniature TVs.

Panel 1. Acronyms in This Paper

BRI	basic rate interface
CCITT	International Telegraph and Telephone Consultative Committee
CIF	common intermediate format
IEC	Interexchange Carrier
ISDN	Integrated Services Digital Network
ISO	International Standards Organization
LEC	Local Exchange Carrier
NTSC	National Television System Committee
QCIF	quarter common intermediate format
SS7	Signaling System 7

Transmission bandwidth is also more available and less expensive than in the past. ISDN is ideally suited to video telephony, and is being increasingly deployed. Full domestic and international ISDN connectivity is not far off.

Local video applications, even for the personal computer (PC), are now available. Thus, it is now technically possible to use the PC in the developing world of ISDN personal video. ISDN terminal products with standards-based digital transmission are already appearing on business desktops. Such terminals are ideal for integrating voice telephony services with personal video.

ISDN Personal Video in Society

In many ways we are already living in a video society. Certainly Cable News Network (CNN) and Music Television (MTV) have become influential in our culture. Video was once a medium where people passively watched professionally produced material. Today, however, daily use of personal video equipment such as camcorders and VCRs has become commonplace.

The proliferation of video technology has accustomed people to seeing themselves on a screen, and has overcome older habits of being "camera shy." No longer is it a novelty to see oneself on a video display. More and more, video is becoming accepted as "the next best thing to being there." In this new environment, it is natural for people to expect video telephones to become more widely available. Indeed, market studies show an emerging demand.

Also, increasing numbers of people are no longer "computer-shy." For many people, the personal computer is becoming almost as familiar as the telephone. High quality personal computer peripherals have encouraged people to expect integrated desktop solutions with entirely new dimensions of functionality, including personal video.

ISDN Personal Video in Business

Business is often quick to adopt new technologies to save money in existing operations, to seize new market opportunities, and to exploit the business potential inherent in the new technology. There is a growing demand for multimedia applications—including video—on the desktop. It is easy to see how video can increase business efficiency by adding a dimension to the ways people communicate and work today. Nor is it difficult to imagine new opportunities for video in the business world. For example, real estate brokers could share pictures on-line with prospective buyers. Banks could add "live" video tellers to remote automatic teller machines (ATMs). And travel agencies could enhance promotional descriptions with live pictures.

The Personal Uses of Video Technology

Even for those with business video experience, personal video will be a new medium. How will this new medium handle the user's needs?

Experience From Video Teleconferencing. Many business people are accustomed to video teleconferencing. The business community has a large installed base of video teleconference rooms and equipment. Some units already "roll around," i.e., are portable or at least transportable. The next logical step is the desktop.

But personal video is more than just a smaller version of what business people are used to. In many ways, personal video has features in common with "room-to-room" video teleconferencing. A primary function of both business and personal video is communication by "talking heads," i.e., people talking face-to-face or even eye-to-eye over a video link. Teleconference users have learned that existing video teleconferencing equipment can function effectively to bring people together. In some ways, the equipment can make communication more spontaneous and more focused than face-to-face meetings. They can also make such meetings less costly in terms of time and travel.

A certain formality accompanies the room-to-room medium. The rooms are usually shared facilities with limited availability. Users must make appointments

(at both ends) and leave their normal office or work space to use the rooms. This alone tends to reduce the sense of spontaneity. Seating and surroundings are often fixed; and equipment and controls often require training for effective use. When the medium and its uses are examined, personal video can be seen as more than miniaturized room teleconferencing. People expect more from desktop personal video service than they get in today's video teleconference room.

Furthermore, personal video is likely to have more diverse uses than in business-related video teleconferencing. The challenge for developers of personal video terminals is to make them adaptable to different situations while keeping them simple and user friendly.

The Personalities of Personal Video. "Face-to-face" mode comes immediately to mind when one thinks of personal video usage. This mode is the straightforward enhancement of a voice telephone call with a real time video picture of the person at the far end. There may be other modes, but this is the common denominator: personal video will always need to be person-to-person.

In the face-to-face mode, the personal video terminal will need to find a place in the user's ordinary workspace. To be accepted, personal video terminals must be designed to lend warmth, friendliness, spontaneity of use, and informality. In its own way, the video terminal must be as available and as easy to use as the voice telephone.

In the role of person-to-person communicator, personal video is an intimate medium. Personal video demands good eye contact. It should be adaptable to different lighting conditions and adjustable to different workspace "moods." Instead of the user going to the video (as in teleconferencing), the video must go to the user, who may wish to use video from different parts of the workspace.

The personal video feature controls should be friendly and obvious. The user needs to be readily aware of both the transmitted and received images. The quality of the video image should be easily adjustable. As an extension to the "mute" function in voice-only communications, the user may also need to invoke "video privacy," which allows temporary suspension of video transmission to the other end.

But personal video has modes beyond person-to-person. Sometimes it is important to see and communicate visually with the people at the other end, either individually or as a group in the teleconferencing room, to share other visual media such as pictures, applications, graphics, text, and documents. Video terminals can address this need by providing an optional graphics mode, as well as a mode for alternative video sources such as videotapes, CD players, and auxiliary cameras. These modes will be especially important for the "business personality" of personal video. They will have their own set of requirements to be successful. It should be easy to change from the face-to-face mode to the graphics or auxiliary video modes and to manipulate the features of each mode.

Residential users will have access to other "personalities" of personal video, and the various types of users in this category must be considered. Full-time work-at-home people or electronic commuters may have the same or similar needs as regular business users, but personal businesses operating from a home might have different needs. At the same time, ordinary personal callers in the residence should not be intimidated by the technology.

Technical Keys to Success

The technology is now in place to develop ISDN video terminals. Two keys to success are at hand:

- Standards being developed by the International Telegraph and Telephone Consultative Committee (CCITT) and International Standards Organization (ISO) have encouraged interoperability between and among products from different vendors.
- Low-cost custom silicon video codec chip sets, such as the AT&T Microelectronics AVP-4000 video-codec chip set, introduced in April 1992, are available.

CCITT Standards and Interoperability

Interoperability of personal video terminals will be achieved by complying with standards for ISDN and for video and audio compression.

In the 1980s, promulgating the CCITT Group 3 standard for facsimile transmission amounted to an open systems standard and was the linchpin for the phenomenal success of facsimile technology. It provided an international interoperability standard for fax communications. The fax industry went from the proprietary domain, where machines from one vendor could not operate with machines from other vendors, to the open systems domain, where all vendors' fax machines could

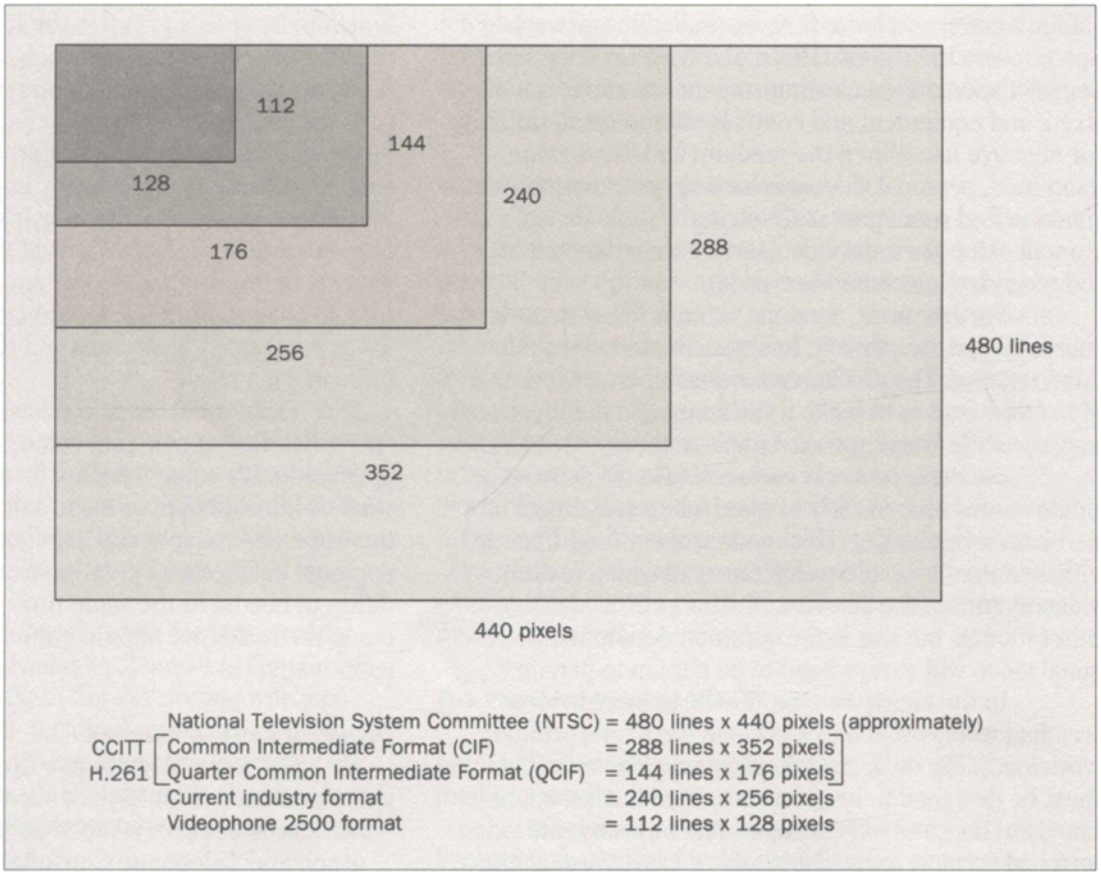

Figure 1. The relative relationship of some video processing formats. A video telephone or video conferencing system built to conform to CCITT standard H.261 must support the quarter common intermediate format (QCIF), 144 lines by 176 pixels, and, optionally, full common intermediate format (CIF), 288 lines by 352 pixels.

interoperate. By 1982, every new fax machine entering the market had Group 3 compatibility.

Similarly, in the video communications area, in December 1990 the CCITT promulgated the H-series family of standards. These included H.261, a digital video compression standard for video telephony and video conferencing. H.261 is expected to have the same effect on the video telephony and the video conferencing business during the 1990s that the Group 3 standard had on the fax business during the 1980s. Before the H.261 standard was adapted, the video telephony and video conferencing business was proprietary. All new video telephony products will probably be H.261-compatible.

The H.261 standard is often referred to as the P×64 standard, where P is an integer between 1 and 30. It covers video codecs operating from 40 kilobits per second (kb/s) to 1.92 megabits per second (Mb/s). Within the H.261 standard, the quarter common intermediate format (QCIF) is specified as the standard video processing format for the video frame. The QCIF video frame consists of 144 lines and 176 pixels per line. The common intermediate format (CIF) is specified as an optional video processing format with 288 lines and 352 pixels per line.

CIF is different from the current domestic de facto standard video processing format in the video conferencing business, and is used by Compression Labs, Inc., PictureTel Corp., and others. That format is 240 lines and 256 pixels per line, about one quarter the resolution of the National Television System Committee (NTSC) 525-line standard. The 525-line broadcast standard produces about 480 lines of analog video output. With interlaced scanning, each of the two fields has about 240 lines. The resulting color resolution is equal to about 330 lines with 440 pixels per line.

Figure 1 illustrates the relative relationship of some video processing formats. Any video telephone or video conferencing system built to conform to H.261

Table I. CCITT H-series standards

Number	Definition
H.221	Frame structure for audiovisual services
H.230	Control and indication for audiovisual services
H.242	Establishing communication between audiovisual terminals
H.261	Video codec for audiovisual services at P×64 kb/s
H.320	Narrow-band visual telephone systems and terminal equipment

must support QCIF (144 lines by 176 pixels) and, optionally, full CIF (288 lines by 352 pixels).

H.261 processing for QCIF video frames operates as follows:

- Each input QCIF frame (up to 30 per second) is digitized and broken into one luminance subframe and two independent chrominance subframes. The luminance (Y) subframe consists of 144 lines, each containing 176 8-bit pixels.
- The chrominance sub-frames are subsampled further. Each contains 72 lines by 88 pixels of red and blue color-difference (R-Y and B-Y) signals.
- The digital compression algorithms—really the heart of the H.261 standard—now reduce the number of bits required to transmit the video. The compression scheme includes elements such as predictive coding, lossy compression by using the discrete cosine transform and quantization (i.e., not all coefficients are transmitted), motion compensation, further digital temporal and spatial filtering, and Huffman-like encoding.

CCITT H.261 is actually part of a suite of standards that includes H.221, H.230, H.242, and H.320. Table I briefly describes the content of each standard. To conform to H.261 and thus be able to interoperate with other vendors' products, a video telephone or video conferencing product must conform to all H-series standards.

An H.261-compliant video telephone or video conferencing system must also support one or more CCITT audio standards. Table II briefly outlines three G-series audio standards. These standards specify the encoding and decoding of audio signals and the compression into various bit rates. It is assumed that virtually every H.261-compliant product will support the G.711 standard for 3 kilohertz (kHz) voice at 64kb/s. While the G.722 standard for 7 kHz voice at 64 kb/s allows better audio quality, it is probable that some vendors will not choose to

Table II. CCITT G-series audio standards

Number	Bit rate	Frequency
G.711	64 kb/s	3 kHz
G.722	64 kb/s	7 kHz (high fidelity)
G.728	16 kb/s	3 kHz

implement the G.722 algorithm. In 112 to 128 kb/s applications, the G.728 standard for 3 kHz voice at 16 kb/s will probably be implemented by many vendors. It will allow more bandwidth to achieve better video quality.

H.261 also includes a mode for high resolution still frame video. In the context of 2-way video communication, this mode will allow alternative transmission of high resolution still frames. This high resolution still frame mode can be used to transmit text, graphics, drawings or photographs.

Custom Silicon Video Codec. The CCITT standards for video telephony have encouraged custom silicon development. The AT&T Microelectronics chip set (AVP-4000) consists of three chips: an encoder, decoder, and systems controller. When used with a digital signal processor, such as the DSP3210, the AVP-4000 chip set provides all the functionality required for video (H.261) and audio (G.711, G.722, and G.728). The chip set also supports the H.xxx suite of standards necessary for full H.261 compliance. The AVP-4000 supports both CIF (288 lines by 352 pixels) and QCIF (144 lines by 176 pixels). This chip set will serve as a strategic core technology within AT&T, and will provide the video codec engine for future video products across AT&T business units. The video codec engine will be the platform for a range of ISDN videotelephones, video PCs, and multimedia workstations.

Why ISDN?

As the name itself implies, ISDN was created as a communications interface to integrate and communicate many types of digital signals. The Basic Rate Interface (BRI), specified in the CCITT I.430 standard, provides the now familiar "2B+D" connectivity. The term *2B* means that two full-duplex 64 kb/s bearer channels are available to carry signals from a source to a destination. These channels allow transmission of digitized voice, circuit-switched or packet-switched data, as well as video signals.

The D-channel is a full-duplex 16 kb/s out-of-band signaling channel. It is always used for call set-up and control, and can also be used to transport user

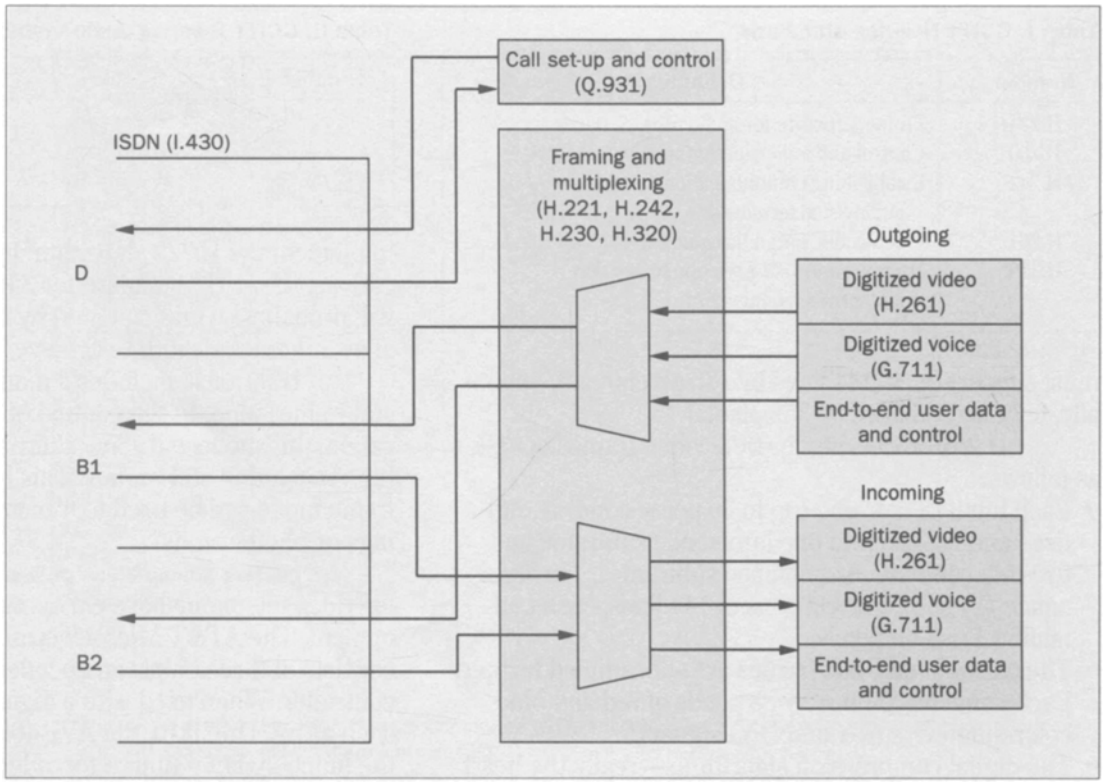

Figure 2. An overview of the video telephony terminal equipment interface to ISDN. The 2B+D capacity of ISDN is depicted as a conduit enclosing the separate parallel B- and D-channels.

packet-data. In video telephony, one or both B-channels are used to transmit the digitized voice and video signals, together with the end-to-end framing and user control signals. The D-channel is used for call setup, and optionally may be used for independent packet data calls.

Interfacing Video Telephony to ISDN. Figure 2 shows an overview of the video telephony terminal equipment interface to ISDN. The 2B+D capacity of ISDN is depicted as a conduit enclosing the separate parallel B- and D-channels. (The BRI, as defined in CCITT I.430, is a serial channel with the B- and D- channels multiplexed together in a 192 kb/s full-duplex channel.)

The terminal equipment uses the D-channel for the initial set-up and teardown of the call through the network. The terminal equipment also multiplexes, demultiplexes, and frames the digital inputs and outputs for the B-channels. The multiplexing and framing of these signals is specified in CCITT H.221. The inputs and outputs are the compressed digital video (H.261), the digital audio (G.711, G.722, or G.728), and the end-to-end user data and control signals.

It is interesting to note that much of the necessary end-to-end user control signaling could theoretically be provided over the D-channel via Q.931 signaling. Instead, the standards facilitate end-to-end control and signaling through in-band data on the B-channels via the H.221 protocol. In this way, ISDN video terminal equipment can interwork with video terminals that are not on ISDN, or are on separate "islands of ISDN," where digital trunking capabilities from end to end or the means to transmit the required ISDN information elements are lacking.

Providing End-to-End Digital Connectivity. ISDN is by definition a digital transmission interface for its terminal endpoints. Thus, each ISDN call is inherently digital, at least between the terminal equipment and the local switch, such as the AT&T Definity® and Legend® customer premises products, and the AT&T 5ESS® switch local central office. Standards-compliant video telephony requires end-to-end digital connectivity. As with data calls, digital switching and trunking must be provided for video calls throughout each leg of the total network connection. However, among the digital facilities provided

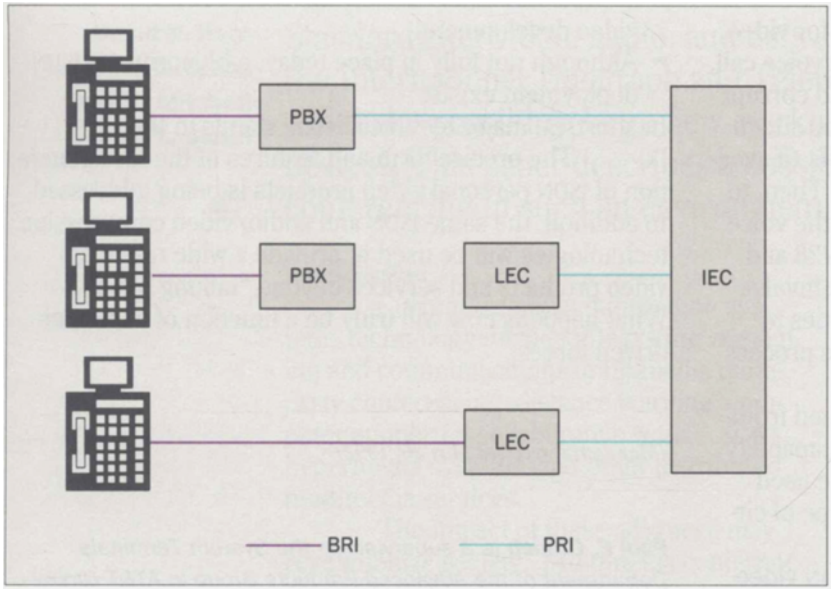

Figure 3. ISDN videotelephony networking alternatives, showing connections to the interexchange carrier (IEC) with and without a private branch exchange (PBX) or local exchange carrier (LEC).

today for data calls, some are not suitable for video calls, and some are more suitable than others.

The best facilities for video connectivity are those providing clear, unrestricted 64 kb/s channels for the ISDN B-channels. These facilities, as well as common channel signaling through Signaling System 7 (SS7), are increasingly provided by the Local Exchange Carriers (LECs). There are also 56 kb/s channels that provide clear, unrestricted connectivity. In the absence of available 64 kb/s channels, video telephony can operate over 56 kb/s facilities. But less bandwidth will necessarily reduce the quality of the video.

Some existing digital facilities cannot be identified as available for video. To establish the required end-to-end circuit-switched data connection, the network must be able to sense the circuit capabilities. The problem is that the network cannot identify some unrestricted data channels. Therefore, the call must be denied. As out-of-band control signaling (SS7) becomes universal, this problem will probably be reduced or completely eliminated.

Signaling and Control for the Transmission System. When a call is made, the ISDN endpoint shows the type of call desired through the bearer capability information element. In the call setup process, one purpose of this element is to specify the transmission facilities needed for successful call completion. Some examples of bearer capabilities mentioned in CCITT I.451 include: speech, G.711 audio, G.722 audio, video, restricted digital information, and unrestricted digital information.

Although a video bearer capability is mentioned in the CCITT standard, it has not yet been adopted as common usage by equipment providers and telephone companies. For the present, the speech bearer capability and the unrestricted digital information bearer capabilities for 64 kb/s and 56 kb/s are all that are generally available for video telephony to use.

For now, it is probable that the following scenario will be used to set up most personal video calls:

- An ordinary voice call is set up to the desired station.
- After the voice call is answered, one or two unrestricted circuit-switched data calls are set up.
- The original voice call is dropped.

There are several reasons for first setting up an ordinary voice call. Sometimes it is not known whether the called station has video capability. Even if it is known, it may not be desirable to add video until the the parties at both ends agree. Either way, voice calls are *always* possible between stations, whereas unrestricted data calls are not available everywhere. Furthermore, voice calls have desirable call handling features that currently are not available with data calls. These are the common features used in establishing contact with the right party: items such as call coverage, voice messaging (e.g., Audix), call forwarding, transfer, and hold.

However, the voice call cannot be used for video. Even if the network uses digital facilities for the voice call, it may insert loss or echo cancellation that would corrupt data transmission. Therefore, when video is to be added to the call, unrestricted circuit-switched data calls (if available) must be established to link the endpoints. Then, to allow as much bandwidth as possible for video, the voice portion of the call may be compressed using G.728 and then conveyed using H.221. Data call setup may involve several attempts to use different bearer capabilities to search out available transmission channels. This process will be improved wherever SS7 is adopted.

Incoming video calls may be distinguished from ordinary incoming data calls through low-layer capability message elements. These message elements are used for end-to-end signaling to show exactly what type of circuit switched data call is being established.

A Scenario to Eliminate Islands of ISDN. ISDN video calls require unrestricted circuit-switched data connectivity, but such connectivity is not yet universally available. At present we have "islands of ISDN" that lack digital trunking capabilities from end to end or the means to transmit the required ISDN information elements. These islands are being bridged and gradually eliminated as SS7 is adopted and as tariffs are established that permit circuit-switched data connectivity.

High quality, high bit-rate digital video service can be provided in several ways (Figure 3). The usual way of achieving circuit-switched data connectivity on IntraLATA calls is to obtain the service from the appropriate LEC. ISDN video endpoints may interface directly to the local central office. Alternatively, ISDN video endpoints on a PBX system may interface to the central office indirectly through T1 lines or ISDN Primary Rate Interface (PRI) service. If equipment or tariff limitations prohibit the service provisioning through the LEC network, then connectivity may be possible through an interexchange carrier (IEC) such as AT&T. For InterLATA video calls, both LECs and IECs are likely to be involved in the needed service, unless the customer has direct connection to an IEC.

Summary

A combination of positive factors is paving the way for ISDN personal video:

- Reasonably priced endpoints are achievable.
- Standards are in place that can promote rapid video development.
- Although not fully in place today, a blueprint for ISDN deployment exists.

In short, no major technical issue stands in the way.

The precise form and features of the first generation of ISDN personal video products is being addressed. In addition, the same ISDN and audio/video compression technologies will be used to provide a wide range of video products and services beyond "talking heads." What happens now will truly be a function of customer-driven forces.

(Manuscript received May 26, 1992)

Paul E. Crouch *is a supervisor in the System Terminals Department of the Advanced Products Group in AT&T Global Business Communications Systems, Holmdel, New Jersey. He coaches a group in the business terminals development area, where his work is currently focused on exploring new video terminal applications for use with ISDN telephone service. He joined AT&T in 1969 with a B.S. in electrical engineering from West Virginia University, Morgantown, an M.S.E. in electrical engineering from Ohio State University, Columbus.*

J. Al Hicks *is supervisor of the Visual Communications Architecture Group in the Architecture Area at AT&T Bell Laboratories, Holmdel, New Jersey. His group has developed and is maintaining AT&T's visual communications architecture that provides a framework for the company's visual communications products and services. He joined AT&T in 1977 with a B.S. in applied psychology from the Georgia Institute of Technology, Atlanta, an M.A. in experimental psychology from the University of Texas at El Paso, and a Ph.D. in human factors psychology from North Carolina State University, Raleigh.*

John J. Jetzt *is a distinguished member of technical staff in the Customer Requirements Department at AT&T Global Business Communications Systems, Holmdel, New Jersey, where he is responsible for systems engineering for business terminals. He joined AT&T in 1963 with a B.E.E. from Manhattan College, New York, and an M.S. in electrical engineering from the Massachusetts Institute of Technology, Cambridge.*